Chapter 20
Ergodic Theory

Laws of large numbers, e.g., for i.i.d. random variables $X_1, X_2, \ldots$, state that the sequence of averages converges a.s. to the expected value, $n^{-1}\sum_{i=1}^{n} X_i \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}[X_1]$. Hence averaging over one realization of many random variables is equivalent to averaging over all possible realizations of one random variable. In the terminology of statistical physics this means that the *time average*, or path (Greek: *odos*) average, equals the *space average*. The "space" in "space average" is the probability space in mathematical terminology, and in physics it is considered the space of admissible states with a certain energy (Greek: *ergon*). Combining the Greek words gives rise to the name *ergodic theory*, which studies laws of large numbers for possibly dependent, but stationary, random variables.

For further reading, see, for example [103] or [88].

20.1 Definitions

Definition 20.1 Let $I \subset \mathbb{R}$ be a set that is closed under addition (for us the important examples are $I = \mathbb{N}_0$, $I = \mathbb{N}$, $I = \mathbb{Z}$, $I = \mathbb{R}$, $I = [0, \infty)$, $I = \mathbb{Z}^d$ and so on). A stochastic process $X = (X_t)_{t\in I}$ is called *stationary* if

$$\mathcal{L}\big[(X_{t+s})_{t\in I}\big] = \mathcal{L}\big[(X_t)_{t\in I}\big] \quad \text{for all } s \in I. \tag{20.1}$$

Remark 20.2 If $I = \mathbb{N}_0$, $I = \mathbb{N}$ or $I = \mathbb{Z}$, then (20.1) is equivalent to

$$\mathcal{L}\big[(X_{n+1})_{n\in I}\big] = \mathcal{L}\big[(X_n)_{n\in I}\big]. \qquad \Diamond$$

Example 20.3

(i) If $X = (X_t)_{t\in I}$ is i.i.d., then X is stationary. If only $\mathbf{P}_{X_t} = \mathbf{P}_{X_0}$ holds for every $t \in I$ (without the independence), then in general X is not stationary. For example, consider $I = \mathbb{N}_0$ and $X_1 = X_2 = X_3 = \ldots$ but $X_0 \neq X_1$. Then X is not stationary.

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_20,

(ii) Let X be a Markov chain with invariant distribution π. If $\mathcal{L}[X_0] = \pi$, then X is stationary.

(iii) Let $(Y_n)_{n\in\mathbb{Z}}$ be i.i.d. real random variables and let $c_1, \ldots, c_k \in \mathbb{R}$. Then

$$X_n := \sum_{l=1}^{k} c_l Y_{n-l}, \quad n \in \mathbb{Z},$$

defines a stationary process X that is called the *moving average* with weights $(c_1, \ldots, c_k)$. In fact, X is stationary if only Y is stationary. $\Diamond$

Lemma 20.4 *If $(X_n)_{n\in\mathbb{N}_0}$ is stationary, then X can be extended to a stationary process $(\widetilde{X}_n)_{n\in\mathbb{Z}}$.*

Proof Let $\widetilde{X}$ be the canonical process on $\Omega = E^{\mathbb{Z}}$. For $n \in \mathbb{N}$, define a probability measure $\widetilde{\mathbf{P}}^{\{-n,-n+1,\ldots\}} \in \mathcal{M}_1(E^{\{-n,-n+1,\ldots\}})$ by

$$\begin{aligned}\widetilde{\mathbf{P}}^{\{-n,-n+1,\ldots\}}&[\widetilde{X}_{-n} \in A_{-n}, \widetilde{X}_{-n+1} \in A_{-n+1}, \ldots] \\ &= \mathbf{P}[X_0 \in A_{-n}, X_1 \in A_{-n+1}, \ldots].\end{aligned}$$

Then $\{-n, -n+1, \ldots\} \uparrow \mathbb{Z}$ and $(\widetilde{\mathbf{P}}^{\{-n,-n+1,\ldots\}}, n \in \mathbb{N})$ is a consistent family. By the Ionescu-Tulcea theorem (Theorem 14.32), the projective limit $\widetilde{\mathbf{P}} := \varprojlim_{n\to\infty} \widetilde{\mathbf{P}}^{\{-n,-n+1,\ldots\}}$ exists. By construction, $\widetilde{X}$ is stationary with respect to $\widetilde{\mathbf{P}}$ and

$$\widetilde{\mathbf{P}} \circ \big((\widetilde{X}_n)_{n\in\mathbb{N}_0}\big)^{-1} = \mathbf{P} \circ \big((X_n)_{n\in\mathbb{N}_0}\big)^{-1}.$$

$\square$

In the following, assume that $(\Omega, \mathcal{A}, \mathbf{P})$ is a probability space and $\tau : \Omega \to \Omega$ is a measurable map.

Definition 20.5 An event $A \in \mathcal{A}$ is called *invariant* if $\tau^{-1}(A) = A$ and *quasi-invariant* if $\mathbb{1}_{\tau^{-1}(A)} = \mathbb{1}_A$ $\mathbf{P}$-a.s. Denote the *σ-algebra of invariant events* by

$$\mathcal{I} = \big\{A \in \mathcal{A} : \tau^{-1}(A) = A\big\}.$$

Recall that a σ-algebra $\mathcal{I}$ is called $\mathbf{P}$-trivial if $\mathbf{P}[A] \in \{0, 1\}$ for every $A \in \mathcal{I}$.

Definition 20.6

(i) τ is called *measure-preserving* if

$$\mathbf{P}\big[\tau^{-1}(A)\big] = \mathbf{P}[A] \quad \text{for all } A \in \mathcal{A}.$$

In this case, $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ is called a *measure-preserving dynamical system*.

(ii) If τ is measure-preserving and $\mathcal{I}$ is $\mathbf{P}$-trivial, then $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ is called *ergodic*.

Lemma 20.7

(i) *A measurable map* $f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is* $\mathcal{I}$*-measurable if and only if* $f \circ \tau = f$.
(ii) $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ *is ergodic if and only if any* $\mathcal{I}$*-measurable* $f : (\Omega, \mathcal{I}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is* $\mathbf{P}$*-almost surely constant.*

Proof (i) The statement is obvious if $f = \mathbb{1}_A$ is an indicator function. The general case, can be inferred by the usual approximation arguments (see Theorem 1.96(i)).

(ii) "$\Longrightarrow$" Assume that $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ is ergodic. Then, for any $c \in \mathbb{R}$, we have $f^{-1}((c, \infty)) \in \mathcal{I}$ and thus $\mathbf{P}[f^{-1}((c, \infty))] \in \{0, 1\}$. We conclude that

$$f = \inf\big\{c \in \mathbb{R} : \mathbf{P}\big[f^{-1}\big((c, \infty)\big)\big] = 0\big\} \quad \mathbf{P}\text{-a.s.}$$

"$\Longleftarrow$" Assume any $\mathcal{I}$-measurable map is $\mathbf{P}$-a.s. constant. If $A \in \mathcal{I}$, then $\mathbb{1}_A$ is $\mathcal{I}$-measurable and hence $\mathbf{P}$-a.s. equals either 0 or 1. Thus $\mathbf{P}[A] \in \{0, 1\}$. □

Example 20.8 Let $n \in \mathbb{N} \setminus \{1\}$, let $\Omega = \mathbb{Z}/(n)$, let $\mathcal{A} = 2^{\Omega}$ and let $\mathbf{P}$ be the uniform distribution on Ω. Let $r \in \{1, \ldots, n\}$ and

$$\tau : \Omega \to \Omega, \quad x \mapsto x + r \pmod n.$$

Then τ is measure-preserving. If $d = \gcd(n, r)$ and

$$A_i = \big\{i, \tau(i), \tau^2(i), \ldots, \tau^{n-1}(i)\big\} = i + \langle r\rangle \quad \text{for } i = 0, \ldots, d-1,$$

then $A_0, \ldots, A_{d-1}$ are the disjoint coset classes of the normal subgroup $\langle r\rangle \trianglelefteq \Omega$. Hence we have $A_i \in \mathcal{I}$ for $i = 0, \ldots, d-1$, and each $A \in \mathcal{I}$ is a union of certain A_i's. Hence we have

$$(\Omega, \mathcal{A}, \mathbf{P}, \tau) \quad \text{is ergodic} \quad \Longleftrightarrow \quad \gcd(r, n) = 1. \qquad \Diamond$$

Example 20.9 (Rotation) Let $\Omega = [0, 1)$, let $\mathcal{A} = \mathcal{B}(\Omega)$ and let $\mathbf{P} = \lambda$ be the Lebesgue measure. Let $r \in (0, 1)$ and $\tau_r(x) = x + r \pmod 1$. Clearly, $(\Omega, \mathcal{A}, \mathbf{P}, \tau_r)$ is a measure-preserving dynamical system. We will show

$$(\Omega, \mathcal{A}, \mathbf{P}, \tau_r) \quad \text{is ergodic} \quad \Longleftrightarrow \quad r \text{ is irrational.}$$

Let $f : [0, 1) \to \mathbb{R}$ be an $\mathcal{I}$-measurable function. Without loss of generality, we assume that f is bounded and thus square integrable. Hence f can be expanded in a Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x} \quad \text{for } \mathbf{P}\text{-a.a. } x.$$

This series converges in L^2, and the sequence of square summable coefficients $(c_n)_{n\in\mathbb{Z}}$ is unique (compare Exercise 7.3.1 with $c_n = (-i/2)a_n + (1/2)b_n$ and

$c_{-n} = (i/2)a_n + (1/2)b_n$ for $n \in \mathbb{N}$ as well as $c_0 = b_0$). Now we compute

$$(f \circ \tau_r)(x) = \sum_{n=-\infty}^{\infty} \left(c_n e^{2\pi i n r}\right) e^{2\pi i n x} \quad \text{a.e.}$$

By Lemma 20.7, f is $\mathcal{I}$-measurable if and only if $f = f \circ \tau_r$; that is, if and only if

$$c_n = c_n e^{2\pi i n r} \quad \text{for all } n \in \mathbb{Z}.$$

If r is irrational, this implies $c_n = 0$ for $n \neq 0$, and thus f is almost surely constant. Therefore, $(\Omega, \mathcal{A}, \mathbf{P}, \tau_r)$ is ergodic.

On the other hand, if r is rational, then there exists some $n \in \mathbb{Z} \setminus \{0\}$ with $e^{2\pi i n r} = e^{-2\pi i n r} = 1$. Hence $x \mapsto e^{2\pi i n x} + e^{-2\pi i n x} = 2\cos(2\pi n x)$ is $\mathcal{I}$-measurable but not a.s. constant. Thus, in this case $(\Omega, \mathcal{A}, \mathbf{P}, \tau_r)$ is not ergodic. ◊

Example 20.10 Let $X = (X_n)_{n \in \mathbb{N}_0}$ be a stochastic process with values in a Polish space E. Without loss of generality, we may assume that X is the canonical process on the probability space $(\Omega, \mathcal{A}, \mathbf{P}) = (E^{\mathbb{N}_0}, \mathcal{B}(E)^{\otimes \mathbb{N}_0}, \mathbf{P})$. Define the *shift* operator

$$\tau : \Omega \to \Omega, \quad (\omega_n)_{n \in \mathbb{N}_0} \mapsto (\omega_{n+1})_{n \in \mathbb{N}_0}.$$

Then $X_n(\omega) = X_0(\tau^n(\omega))$. Hence X is stationary if and only if $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ is a measure-preserving dynamical system. ◊

Definition 20.11 The stochastic process X (from Example 20.10) is called *ergodic* if $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ is ergodic.

Example 20.12 Let $(X_n)_{n \in \mathbb{N}_0}$ be i.i.d. and let $X_n(\omega) = X_0(\tau^n(\omega))$. If $A \in \mathcal{I}$, then, for every $n \in \mathbb{N}$,

$$A = \tau^{-n}(A) = \left\{\omega : \tau^n(\omega) \in A\right\} \in \sigma(X_n, X_{n+1}, \ldots).$$

Hence, if we let $\mathcal{T}$ be the tail σ-algebra of $(X_n)_{n \in \mathbb{N}}$ (see Definition 2.34), then

$$\mathcal{I} \subset \mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \ldots).$$

By Kolmogorov's 0–1 law (Theorem 2.37), $\mathcal{T}$ is $\mathbf{P}$-trivial. Hence $\mathcal{I}$ is also $\mathbf{P}$-trivial and therefore $(X_n)_{n \in \mathbb{N}_0}$ is ergodic. ◊

Exercise 20.1.1 Let G be a finite group of measure-preserving measurable maps on $(\Omega, \mathcal{A}, \mathbf{P})$ and let $\mathcal{A}_0 := \{A \in \mathcal{A} : g(A) = A \text{ for all } g \in G\}$.

Show that, for every $X \in \mathcal{L}^1(\mathbf{P})$, we have

$$\mathbf{E}[X \mid \mathcal{A}_0] = \frac{1}{\#G} \sum_{g \in G} X \circ g.$$

20.2 Ergodic Theorems

In this section, $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ always denotes a measure-preserving dynamical system. Further, let $f : \Omega \to \mathbb{R}$ be measurable and

$$X_n(\omega) = f \circ \tau^n(\omega) \quad \text{for all } n \in \mathbb{N}_0.$$

Hence $X = (X_n)_{n\in\mathbb{N}_0}$ is a stationary real-valued stochastic process. Let

$$S_n = \sum_{k=0}^{n-1} X_k$$

denote the nth partial sum. Ergodic theorems are laws of large numbers for $(S_n)_{n\in\mathbb{N}}$. We start with a preliminary lemma.

Lemma 20.13 (Hopf's maximal-ergodic lemma) *Let $X_0 \in \mathcal{L}^1(\mathbf{P})$. Define $M_n = \max\{0, S_1, \ldots, S_n\}$, $n \in \mathbb{N}$. Then*

$$\mathbf{E}[X_0 \mathbb{1}_{\{M_n>0\}}] \geq 0 \quad \textit{for every } n \in \mathbb{N}.$$

Proof For $k \leq n$, we have $M_n(\tau(\omega)) \geq S_k(\tau(\omega))$. Hence

$$X_0 + M_n \circ \tau \geq X_0 + S_k \circ \tau = S_{k+1}.$$

Thus $X_0 \geq S_{k+1} - M_n \circ \tau$ for $k = 1, \ldots, n$. Manifestly, $S_1 = X_0$ and $M_n \circ \tau \geq 0$ and hence also (for $k = 0$) $X_0 \geq S_1 - M_n \circ \tau$. Therefore,

$$X_0 \geq \max\{S_1, \ldots, S_n\} - M_n \circ \tau. \tag{20.2}$$

Furthermore, we have

$$\{M_n > 0\}^c \subset \{M_n = 0\} \cap \{M_n \circ \tau \geq 0\} \subset \{M_n - M_n \circ \tau \leq 0\}. \tag{20.3}$$

By (20.2) and (20.3), and since τ is measure-preserving, we conclude that

$$\begin{aligned}
\mathbf{E}[X_0 \mathbb{1}_{\{M_n>0\}}] &\geq \mathbf{E}\big[\big(\max\{S_1, \ldots, S_n\} - M_n \circ \tau\big)\mathbb{1}_{\{M_n>0\}}\big] \\
&= \mathbf{E}\big[(M_n - M_n \circ \tau)\mathbb{1}_{\{M_n>0\}}\big] \\
&\geq \mathbf{E}[M_n - M_n \circ \tau] = \mathbf{E}[M_n] - \mathbf{E}[M_n] = 0.
\end{aligned}$$

□

Theorem 20.14 (Individual ergodic theorem, Birkhoff (1931), [16]) *Let* $f = X_0 \in \mathcal{L}^1(\mathbf{P})$. *Then*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k = \frac{1}{n}\sum_{k=0}^{n-1} f \circ \tau^k \overset{n\to\infty}{\longrightarrow} \mathbf{E}[X_0 \,|\, \mathcal{I}] \quad \mathbf{P}\text{-}a.s.$$

In particular, if τ *is ergodic, then*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \overset{n\to\infty}{\longrightarrow} \mathbf{E}[X_0] \quad \mathbf{P}\text{-}a.s.$$

Proof If τ is ergodic, then $\mathbf{E}[X_0 \,|\, \mathcal{I}] = \mathbf{E}[X_0]$ and the supplement is a consequence of the first statement.

Consider now the general case. By Lemma 20.7, we have $\mathbf{E}[X_0 \,|\, \mathcal{I}] \circ \tau = \mathbf{E}[X_0 \,|\, \mathcal{I}]$ $\mathbf{P}$-a.s. Hence, by passing to $\widetilde{X}_n := X_n - \mathbf{E}[X_0 \,|\, \mathcal{I}]$, without loss of generality, we can assume $\mathbf{E}[X_0 \,|\, \mathcal{I}] = 0$. Define

$$Z := \limsup_{n\to\infty} \frac{1}{n} S_n.$$

Let $\varepsilon > 0$ and $F := \{Z > \varepsilon\}$. We have to show that $\mathbf{P}[F] = 0$. From this we infer $\mathbf{P}[Z > 0] = 0$ and similarly (with $-X$ instead of X) also $\liminf_{n\to\infty} \frac{1}{n} S_n \geq 0$ almost surely. Hence $\frac{1}{n} S_n \overset{n\to\infty}{\longrightarrow} 0$ a.s.

Evidently, $Z \circ \tau = Z$; hence $F \in \mathcal{I}$. Define

$$X_n^\varepsilon := (X_n - \varepsilon)\mathbb{1}_F, \qquad S_n^\varepsilon := X_0^\varepsilon + \ldots + X_{n-1}^\varepsilon,$$
$$M_n^\varepsilon := \max\{0, S_1^\varepsilon, \ldots, S_n^\varepsilon\}, \qquad F_n := \{M_n^\varepsilon > 0\}.$$

Then $F_1 \subset F_2 \subset \ldots$ and

$$\bigcup_{n=1}^{\infty} F_n = \left\{\sup_{k\in\mathbb{N}} \frac{1}{k} S_k^\varepsilon > 0\right\} = \left\{\sup_{k\in\mathbb{N}} \frac{1}{k} S_k > \varepsilon\right\} \cap F = F,$$

hence $F_n \uparrow F$. Dominated convergence yields $\mathbf{E}[X_0^\varepsilon \mathbb{1}_{F_n}] \overset{n\to\infty}{\longrightarrow} \mathbf{E}[X_0^\varepsilon]$.

By the maximal-ergodic lemma (applied to X^ε), we have $\mathbf{E}[X_0^\varepsilon \mathbb{1}_{F_n}] \geq 0$; hence

$$\begin{aligned} 0 \leq \mathbf{E}\big[X_0^\varepsilon\big] &= \mathbf{E}\big[(X_0 - \varepsilon)\mathbb{1}_F\big] \\ &= \mathbf{E}\big[\mathbf{E}[X_0 \,|\, \mathcal{I}]\mathbb{1}_F\big] - \varepsilon\mathbf{P}[F] = -\varepsilon\mathbf{P}[F]. \end{aligned}$$

We conclude that $\mathbf{P}[F] = 0$. □

As a consequence, we obtain the statistical ergodic theorem, or L^p-ergodic theorem, that was found by von Neumann in 1931 right before Birkhoff proved his ergodic theorem, but was published only later in [122]. Before we formulate it, we state one more lemma.

Lemma 20.15 *Let $p \geq 1$ and let $X_0, X_1, \ldots$ be identically distributed, real random variables with $\mathbf{E}[|X_0|^p] < \infty$. Define $Y_n := |\frac{1}{n}\sum_{k=0}^{n-1} X_k|^p$ for $n \in \mathbb{N}$. Then $(Y_n)_{n\in\mathbb{N}}$ is uniformly integrable.*

Proof Evidently, the singleton $\{|X_0|^p\}$ is uniformly integrable. Hence, by Theorem 6.19, there exists a monotone increasing convex map $f : [0, \infty) \to [0, \infty)$ with $\frac{f(x)}{x} \to \infty$ for $x \to \infty$ and $C := \mathbf{E}[f(|X_0|^p)] < \infty$. Again, by Theorem 6.19, it is enough to show that $\mathbf{E}[f(Y_n)] \leq C$ for every $n \in \mathbb{N}$. By Jensen's inequality (for $x \mapsto |x|^p$), we have

$$Y_n \leq \frac{1}{n}\sum_{k=0}^{n-1} |X_k|^p.$$

Again, by Jensen's inequality (now applied to f), we get that

$$f(Y_n) \leq f\left(\frac{1}{n}\sum_{k=0}^{n-1} |X_k|^p\right) \leq \frac{1}{n}\sum_{k=0}^{n-1} f\big(|X_k|^p\big).$$

Hence $\mathbf{E}[f(Y_n)] \leq \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{E}[f(|X_k|^p)] = C$. □

Theorem 20.16 (L^p-ergodic theorem, von Neumann (1931)) *Let $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ be a measure-preserving dynamical system, $p \geq 1$, $X_0 \in \mathcal{L}^p(\mathbf{P})$ and $X_n = X_0 \circ \tau^n$. Then*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}[X_0 \,|\, \mathcal{I}] \quad \text{in } L^p(\mathbf{P}).$$

In particular, if τ is ergodic, then $\frac{1}{n}\sum_{k=0}^{n-1} X_k \stackrel{n\to\infty}{\longrightarrow} \mathbf{E}[X_0]$ in $L^p(\mathbf{P})$.

Proof Define

$$Y_n := \left|\frac{1}{n}\sum_{k=0}^{n-1} X_k - \mathbf{E}[X_0 \,|\, \mathcal{I}]\right|^p \quad \text{for every } n \in \mathbb{N}.$$

By Lemma 20.15, $(Y_n)_{n\in\mathbb{N}}$ is uniformly integrable, and by Birkhoff's ergodic theorem, we have $Y_n \stackrel{n\to\infty}{\longrightarrow} 0$ almost surely. By Theorem 6.25, we thus have $\lim_{n\to\infty} \mathbf{E}[Y_n] = 0$.

If τ is ergodic, then $\mathbf{E}[X_0 \,|\, \mathcal{I}] = \mathbf{E}[X_0]$. □

20.3 Examples

Example 20.17 Let $(X, (\mathbf{P}_x)_{x\in E})$ be a positive recurrent, irreducible Markov chain on the countable space E. Let π be the invariant distribution of X. Then $\pi(\{x\}) > 0$ for every $x \in E$. Define $\mathbf{P}_\pi = \sum_{x\in E} \pi(\{x\})\mathbf{P}_x$. Then X is stationary on $(\Omega, \mathcal{A}, \mathbf{P}_\pi)$. Denote the shift by τ; that is, $X_n = X_0 \circ \tau^n$.

Now let $A \in \mathcal{I}$ be invariant. Then $A \in \mathcal{T} = \bigcap_{n=1}^\infty \sigma(X_n, X_{n+1}, \ldots)$. By the strong Markov property, for every finite stopping time σ (recall that $\mathcal{F}_\sigma$ is the σ-algebra of the σ-past),

$$\mathbf{P}_\pi[X \in A \mid \mathcal{F}_\sigma] = \mathbf{P}_{X_\sigma}[X \in A]. \tag{20.4}$$

Indeed, we have $\{X \in A\} = \{X \in \tau^{-n}(A)\} = \{(X_n, X_{n+1}, \ldots) \in A\}$. For $B \in \mathcal{F}_\sigma$, using the Markov property (in the third line), we get

$$\begin{aligned}
\mathbf{E}_\pi[\mathbb{1}_{\{X\in B\}}\mathbb{1}_{\{X\in A\}}] &= \sum_{n=0}^\infty \sum_{x\in E} \mathbf{P}_\pi[X \in B, \sigma = n, X_n = x, X \in A] \\
&= \sum_{n=0}^\infty \sum_{x\in E} \mathbf{P}_\pi\big[X \in B, \sigma = n, X_n = x, X \circ \tau^n \in A\big] \\
&= \sum_{n=0}^\infty \sum_{x\in E} \mathbf{P}_\pi[X \in B, \sigma = n, X_n = x]\mathbf{P}_x[X \in A] \\
&= \mathbf{E}_\pi\big[\mathbb{1}_{\{X\in B\}}\mathbf{P}_{X_\sigma}[X \in A]\big].
\end{aligned}$$

In particular, if $x \in E$ and $\sigma_x = \inf\{n \in \mathbb{N}_0 : X_n = x\}$, then $\sigma_x < \infty$ since X is recurrent and irreducible. By (20.4), we conclude that, for every $x \in E$,

$$\mathbf{P}_\pi[X \in A] = \mathbf{E}_\pi\big[\mathbf{P}_x[X \in A]\big] = \mathbf{P}_x[X \in A].$$

Hence $\mathbf{P}_{X_n}[X \in A] = \mathbf{P}_\pi[X \in A]$ almost surely and thus (with $\sigma = n$ in (20.4))

$$\mathbf{P}_\pi[X \in A \mid X_0, \ldots, X_n] = \mathbf{P}_{X_n}[X \in A] = \mathbf{P}_\pi[X \in A].$$

Now $A \in \mathcal{I} \subset \sigma(X_1, X_2, \ldots)$; hence

$$\mathbf{P}_\pi[X \in A \mid X_0, \ldots, X_n] \stackrel{n\to\infty}{\longrightarrow} \mathbf{P}_\pi\big[X \in A | \sigma(X_0, X_1, \ldots)\big] = \mathbb{1}_{\{X\in A\}}.$$

This implies $\mathbf{P}_\pi[X \in A] \in \{0, 1\}$. Hence X is ergodic.

Birkhoff's ergodic theorem now implies that, for every $x \in E$,

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}} \stackrel{n\to\infty}{\longrightarrow} \pi\big(\{x\}\big) \quad \mathbf{P}_\pi\text{-a.s.}$$

In this sense, $\pi(\{x\})$ is the average time X spends in x in the long run. ◊

Example 20.18 Let P and Q be probability measures on the measurable space $(\Omega, \mathcal{A})$, and let $(\Omega, \mathcal{A}, P, \tau)$ and $(\Omega, \mathcal{A}, Q, \tau)$ be ergodic. Then either $P = Q$ or $P \perp Q$. Indeed, if $P \neq Q$, then there exists an f with $|f| \leq 1$ and $\int f\,dP \neq \int f\,dQ$. However, by Birkhoff's ergodic theorem,

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ \tau^k \stackrel{n\to\infty}{\longrightarrow} \begin{cases} \int f\,dP & P\text{-a.s.,} \\ \int f\,dQ & Q\text{-a.s.} \end{cases}$$

If we define $A := \{\frac{1}{n}\sum_{k=0}^{n-1} f \circ \tau^k \stackrel{n\to\infty}{\longrightarrow} \int f\,dP\}$, then $P(A) = 1$ and $Q(A) = 0$. Thus $P \perp Q$. ◊

Exercise 20.3.1 Let $(\Omega, \mathcal{A})$ be a measurable space and let $\tau : \Omega \to \Omega$ be a measurable map.

(i) Show that the set $\mathcal{M} := \{\mu \in \mathcal{M}_1(\Omega) : \mu \circ \tau^{-1} = \mu\}$ of τ-invariant measures is convex.
(ii) An element μ of $\mathcal{M}$ is called *extremal* if $\mu = \lambda\mu_1 + (1 - \lambda)\mu_2$ for some $\mu_1, \mu_2 \in \mathcal{M}$ and $\lambda \in (0, 1)$ implies $\mu = \mu_1 = \mu_2$. Show that $\mu \in \mathcal{M}$ is extremal if and only if τ is ergodic with respect to μ.

Exercise 20.3.2 Let $p = 2, 3, 5, 6, 7, 10, \ldots$ be square-free (that is, there is no number $r = 2, 3, 4, \ldots$, whose square is a divisor of p) and let $q \in \{2, 3, \ldots, p - 1\}$. For every $n \in \mathbb{N}$, let a_n be the leading digit of the p-adic expansion of q^n.

Show the following version of Benford's law: For every $d \in \{1, \ldots, p - 1\}$,

$$\frac{1}{n}\#\{i \leq n : a_i = d\} \stackrel{n\to\infty}{\longrightarrow} \frac{\log(d + 1) - \log(d)}{\log(p)}.$$

20.4 Application: Recurrence of Random Walks

Let $(X_n)_{n\in\mathbb{N}}$ be a stationary process with values in $\mathbb{R}^d$. Define $S_n := \sum_{k=1}^{n} X_k$ for $n \in \mathbb{N}_0$. Further, let

$$R_n = \#\{S_1, \ldots, S_n\}$$

denote the *range* of S; that is, the number of distinct points visited by S up to time n. Finally, let $A := \{S_n \neq 0 \text{ for every } n \in \mathbb{N}\}$ be the event of an "escape" from 0.

Theorem 20.19 *We have* $\lim_{n\to\infty} \frac{1}{n} R_n = \mathbf{P}[A \mid \mathcal{I}]$ *almost surely.*

Proof Let X be the canonical process on $(\Omega, \mathcal{A}, \mathbf{P}) = ((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathbf{P})$ and let $\tau : \Omega \to \Omega$ be the shift; that is, $X_n = X_0 \circ \tau^n$.

Evidently,

$$\begin{aligned}R_n &= \#\big\{k \le n : S_l \ne S_k \text{ for all } l \in \{k+1,\ldots,n\}\big\}\\ &\ge \#\{k \le n : S_l \ne S_k \text{ for all } l > k\}\\ &= \sum_{k=1}^{n} \mathbb{1}_A \circ \tau^k.\end{aligned}$$

Birkhoff's ergodic theorem yields

$$\liminf_{n\to\infty} \frac{1}{n} R_n \ge \mathbf{P}[A \,|\, \mathcal{I}] \quad \text{a.s.} \tag{20.5}$$

For the converse inequality, consider $A_m = \{S_l \ne 0 \text{ for } l = 1,\ldots,m\}$. Then, for every $n \ge m$,

$$\begin{aligned}R_n &\le m + \#\big\{k \le n-m : S_l \ne S_k \text{ for all } l \in \{k+1,\ldots,n\}\big\}\\ &\le m + \#\big\{k \le n-m : S_l \ne S_k \text{ for all } l \in \{k+1,\ldots,k+m\}\big\}\\ &= m + \sum_{k=1}^{n-m} \mathbb{1}_{A_m} \circ \tau^k.\end{aligned}$$

Again, by the ergodic theorem,

$$\limsup_{n\to\infty} \frac{1}{n} R_n \le \mathbf{P}[A_m \,|\, \mathcal{I}] \quad \text{a.s.} \tag{20.6}$$

Since $A_m \downarrow A$ and $\mathbf{P}[A_m \mid \mathcal{I}] \overset{n\to\infty}{\longrightarrow} \mathbf{P}[A \mid \mathcal{I}]$ almost surely (by Theorem 8.14(viii)), the claim follows from (20.5) and (20.6). □

Theorem 20.20 *Let $X = (X_n)_{n\in\mathbb{N}}$ be an integer-valued, integrable, stationary process with the property $\mathbf{E}[X_1 \mid \mathcal{I}] = 0$ a.s. Let $S_n = X_1 + \ldots + X_n$, $n \in \mathbb{N}$. Then*

$$\mathbf{P}[S_n = 0 \textit{ for infinitely many } n \in \mathbb{N}] = 1.$$

In particular, a random walk on $\mathbb{Z}$ with centered increments is recurrent (Chung–Fuchs theorem, compare Theorem 17.40).

Proof Define $A = \{S_n \ne 0 \text{ for all } n \in \mathbb{N}\}$.

Step 1. We show $\mathbf{P}[A] = 0$. (If X is i.i.d., then S is a Markov chain, and this implies immediately that 0 is recurrent. Only for the more general case of stationary X do we need an additional argument.) By the ergodic theorem, we have

$\frac{1}{n}S_n \overset{n\to\infty}{\longrightarrow} \mathbf{E}[X_1 \,|\, \mathcal{I}] = 0$ a.s. Thus, for every $m \in \mathbb{N}$,

$$\limsup_{n\to\infty}\left(\frac{1}{n}\max_{k=1,\ldots,n}|S_k|\right) = \limsup_{n\to\infty}\left(\frac{1}{n}\max_{k=m,\ldots,n}|S_k|\right) \le \max_{k\ge m}\frac{|S_k|}{k} \overset{m\to\infty}{\longrightarrow} 0.$$

Therefore,

$$\lim_{n\to\infty}\left(\frac{1}{n}\max_{k=1,\ldots,n} S_k\right) = \lim_{n\to\infty}\left(\frac{1}{n}\min_{k=1,\ldots,n} S_k\right) = 0.$$

Now $R_n \le 1 + (\max_{k=1,\ldots,n} S_k) - (\min_{k=1,\ldots,n} S_k)$; hence $\frac{1}{n}R_n \overset{n\to\infty}{\longrightarrow} 0$. By Theorem 20.19, this implies $\mathbf{P}[A] = 0$.

Step 2. Define $\sigma_n := \inf\{m \in \mathbb{N} : S_{m+n} = S_n\}$, $B_n := \{\sigma_n < \infty\}$ for $n \in \mathbb{N}_0$ and $B := \bigcap_{n=0}^{\infty} B_n$.

Since $\{\sigma_0 = \infty\} = A$, we have $\mathbf{P}[\sigma_0 < \infty] = 1$. By stationarity, $\mathbf{P}[\sigma_n < \infty] = 1$ for every $n \in \mathbb{N}_0$; hence $\mathbf{P}[B] = 1$.

Let $\tau_0 = 0$ and inductively define $\tau_{n+1} = \tau_n + \sigma_{\tau_n}$ for $n \in \mathbb{N}_0$. Then τ_n is the time of the nth return of S to 0. On B we have $\tau_n < \infty$ for every $n \in \mathbb{N}_0$ and hence

$$\mathbf{P}[S_n = 0 \text{ infinitely often}] = \mathbf{P}[\tau_n < \infty \text{ for all } n \in \mathbb{N}] \ge \mathbf{P}[B] = 1. \qquad \square$$

If in Theorem 20.20 the random variables X_n are not integer-valued, then there is no hope that $S_n = 0$ for any $n \in \mathbb{N}$ with positive probability. On the other hand, in this case, there is also some kind of recurrence property, namely $S_n/n \overset{n\to\infty}{\longrightarrow} 0$ almost surely by the ergodic theorem. Note, however, that this does not exclude the possibility that $S_n \overset{n\to\infty}{\longrightarrow} \infty$ with positive probability; for instance, if S_n grows like $\sqrt{n}$. The next theorem shows that if the X_n are integrable, then the process of partial sums can go to infinity only with a linear speed.

Theorem 20.21 *Let $(X_n)_{n\in\mathbb{N}}$ be an integrable ergodic process and define $S_n = X_1 + \ldots + X_n$ for $n \in \mathbb{N}_0$. Then the following statements are equivalent.*

(i) $S_n \overset{n\to\infty}{\longrightarrow} \infty$ *almost surely.*
(ii) $\mathbf{P}[S_n \overset{n\to\infty}{\longrightarrow} \infty] > 0$.
(iii) $\lim_{n\to\infty} \frac{S_n}{n} = \mathbf{E}[X_1] > 0$ *almost surely.*

If the random variables $X_1, X_2, \ldots$ are i.i.d. with $\mathbf{E}[X_1] = 0$ and $\mathbf{P}[X_1 = 0] < 1$, then $\liminf_{n\to\infty} S_n = -\infty$ and $\limsup_{n\to\infty} S_n = \infty$ almost surely.

Proof "(i) $\Longleftrightarrow$ (ii)" Clearly, $\{S_n \overset{n\to\infty}{\longrightarrow} \infty\}$ is an invariant event and thus has probability either 0 or 1.

"(iii)$\Longrightarrow$(i)" This is trivial.

"(i)$\Longrightarrow$(iii)" The equality follows by the individual ergodic theorem. Hence, it is enough to show that $\liminf_{n\to\infty} S_n/n > 0$ almost surely.

For $n \in \mathbb{N}_0$ and $\varepsilon > 0$, let

$$A_n^\varepsilon := \{S_m > S_n + \varepsilon \text{ for all } m \geq n+1\}.$$

Let $S^- := \inf\{S_n : n \in \mathbb{N}_0\}$. By assumption (i), we have $S^- > -\infty$ almost surely and $\tau := \sup\{n \in \mathbb{N}_0 : S_n = S^-\}$ is finite almost surely. Hence there is an $N \in \mathbb{N}$ with $\mathbf{P}[\tau < N] \geq \frac{1}{2}$. Therefore,

$$\mathbf{P}\left[\bigcup_{n=0}^{N-1} A_n^0\right] = \mathbf{P}[\tau < N] \geq \frac{1}{2}.$$

Since $A_n^\varepsilon \uparrow A_n^0$ for $\varepsilon \downarrow 0$, there is an $\varepsilon > 0$ with $p := \mathbf{P}[A_0^\varepsilon] \geq \frac{1}{4N} > 0$. As $(X_n)_{n\in\mathbb{N}}$ is ergodic, $(\mathbb{1}_{A_n^\varepsilon})_{n\in\mathbb{N}_0}$ is also ergodic. By the individual ergodic theorem, we conclude that $\frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{A_n^\varepsilon} \overset{n\to\infty}{\longrightarrow} p$ almost surely. Hence there exists an $n_0 = n_0(\omega)$ such that $\sum_{i=0}^{n-1} \mathbb{1}_{A_n^\varepsilon} \geq \frac{pn}{2}$ for all $n \geq n_0$. This implies $S_n \geq \frac{pn\varepsilon}{2}$ for $n \geq n_0$ and hence $\liminf_{n\to\infty} S_n/n \geq \frac{pn\varepsilon}{2} > 0$.

The additional statement follows since $\liminf S_n$ and $\limsup S_n$ cannot assume any finite value and are thus measurable with respect to the tail σ-algebra, which implies that they are constantly $-\infty$ or $+\infty$. By what we have shown, we can exclude $S_n \overset{n\to\infty}{\longrightarrow} \infty$; hence we have $\liminf_{n\to\infty} S_n = -\infty$. Similarly, we get $\limsup_{n\to\infty} S_n = \infty$. □

Remark 20.22 It can be shown that Theorem 20.21 holds also without the assumption that the X_n are integrable. See [94]. ◊

20.5 Mixing

Ergodicity provides a weak notion of "independence" or "mixing". At the other end of the scale, the strongest notion is "i.i.d.". Here we are concerned with notions of mixing that lie between these two.

In the following, we always assume that $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ is a measure-preserving dynamical system and that $X_n := X_0 \circ \tau^n$. We start with a simple observation.

Theorem 20.23 $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ *is ergodic if and only if, for all* $A, B \in \mathcal{A}$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{P}\big[A \cap \tau^{-k}(B)\big] = \mathbf{P}[A]\mathbf{P}[B]. \tag{20.7}$$

Proof "$\Longrightarrow$" Let $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ be ergodic. Define

$$Y_n := \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\tau^{-k}(B)} = \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_B \circ \tau^k.$$

By Birkhoff's ergodic theorem, we have $Y_n \overset{n\to\infty}{\longrightarrow} \mathbf{P}[B]$ almost surely. Hence $Y_n \mathbb{1}_A \overset{n\to\infty}{\longrightarrow} \mathbb{1}_A \mathbf{P}[B]$ almost surely. Dominated convergence yields

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbf{P}\big[A \cap \tau^{-k}(B)\big] = \mathbf{E}[Y_n \mathbb{1}_A] \overset{n\to\infty}{\longrightarrow} \mathbf{E}\big[\mathbb{1}_A \mathbf{P}[B]\big] = \mathbf{P}[A]\mathbf{P}[B].$$

"$\Longleftarrow$" Now assume that (20.7) holds. Let $A \in \mathcal{I}$ (recall that $\mathcal{I}$ is the invariant σ-algebra) and $B = A$. Evidently, $A \cap \tau^{-k}(A) = A$ for every $k \in \mathbb{N}_0$. Hence, by (20.7),

$$\mathbf{P}[A] = \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{P}\big[A \cap \tau^{-k}(A)\big] \overset{n\to\infty}{\longrightarrow} \mathbf{P}[A]^2.$$

Thus $\mathbf{P}[A] \in \{0, 1\}$; hence $\mathcal{I}$ is trivial and therefore τ is ergodic. □

We consider a strengthening of (20.7).

Definition 20.24 A measure-preserving dynamical system $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ is called *mixing* if

$$\lim_{n\to\infty} \mathbf{P}\big[A \cap \tau^{-n}(B)\big] = \mathbf{P}[A]\mathbf{P}[B] \quad \text{for all } A, B \in \mathcal{A}. \tag{20.8}$$

Remark 20.25 Sometimes the mixing property of (20.8) is called *strongly mixing*, in contrast with a *weakly mixing* system $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$, for which we require only

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \Big|\mathbf{P}\big[A \cap \tau^{-i}(B)\big] - \mathbf{P}[A]\mathbf{P}[B]\Big| = 0 \quad \text{for all } A, B \in \mathcal{A}.$$

"Strongly mixing" implies "weakly mixing" (see Exercise 20.5.1). On the other hand, there exist weakly mixing systems that are not strongly mixing (see [81]). ◊

Example 20.26 Let $I = \mathbb{N}_0$ or $I = \mathbb{Z}$, and let $(X_n)_{n\in I}$ be an i.i.d. sequence with values in the measurable space $(E, \mathcal{E})$. Hence τ is the shift on the product space $\Omega = E^I, \mathbf{P} = (\mathbf{P}_{X_0})^{\otimes I}$. Let $A, B \in \mathcal{E}^{\otimes I}$. For every $\varepsilon > 0$, there exist events A^ε and B^ε that depend on only finitely many coordinates and such that $\mathbf{P}[A \triangle A^\varepsilon] < \varepsilon$ and $\mathbf{P}[B \triangle B^\varepsilon] < \varepsilon$. Clearly, $\mathbf{P}[\tau^{-n}(A \triangle A^\varepsilon)] < \varepsilon$ and $\mathbf{P}[\tau^{-n}(B \triangle B^\varepsilon)] < \varepsilon$ for every $n \in \mathbb{Z}$. For sufficiently large $|n|$, the sets A^ε and $\tau^{-n}(B^\varepsilon)$ depend on different coordinates and are thus independent. This implies

$$\begin{aligned}\limsup_{|n|\to\infty} &\Big|\mathbf{P}\big[A \cap \tau^{-n}(B)\big] - \mathbf{P}[A]\mathbf{P}[B]\Big| \\ &\le \limsup_{|n|\to\infty} \Big|\mathbf{P}\big[A^\varepsilon \cap \tau^{-n}\big(B^\varepsilon\big)\big] - \mathbf{P}\big[A^\varepsilon\big]\mathbf{P}\big[B^\varepsilon\big]\Big| + 4\varepsilon = 4\varepsilon.\end{aligned}$$

Hence τ is mixing. Letting $A = B \in \mathcal{I}$, we obtain the 0–1 law for invariant events: $\mathbf{P}[A] \in \{0, 1\}$. ◊

Remark 20.27 Clearly, (20.8) implies (20.7) and hence "mixing" implies "ergodic". The converse implication is false. ◊

Example 20.28 Let $\Omega = [0, 1)$, $\mathcal{A} = \mathcal{B}([0, 1))$ and let $\mathbf{P} = \lambda$ be the Lebesgue measure on $([0, 1), \mathcal{B}([0, 1)))$. For $r \in [0, 1)$, define $\tau_r : \Omega \to \Omega$ by

$$\tau_r(x) = x + r - \lfloor x + r \rfloor = x + r \pmod 1.$$

If r is irrational, then τ_r is ergodic (Example 20.9). However, τ_r is not mixing: Since r is irrational, there exists a sequence $k_n \uparrow \infty$ such that

$$\tau_r^{k_n}(0) \in \left(\frac{1}{4}, \frac{3}{4}\right) \quad \text{for } n \in \mathbb{N}.$$

Hence, for $A = [0, \frac{1}{4}]$, we have $A \cap \tau_r^{-k_n}(A) = \emptyset$. Therefore,

$$\liminf_{n\to\infty} \mathbf{P}\big[A \cap \tau_r^{-n}(A)\big] = 0 \neq \frac{1}{16} = \mathbf{P}[A]^2.$$ ◊

Theorem 20.29 *Let X be an irreducible, positive recurrent Markov chain on the countable space E and let π be its invariant distribution. Let $\mathbf{P}_\pi = \sum_{x\in E} \pi(x)\mathbf{P}_x$. Then:*

(i) *X is ergodic (on $(\Omega, \mathcal{A}, \mathbf{P}_\pi)$).*
(ii) *X is mixing if and only if X is aperiodic.*

Proof (i) This has been shown already in Example 20.17.

(ii) As X is irreducible, by Theorem 17.51, we have $\pi(\{x\}) > 0$ for every $x \in E$.

"$\Longrightarrow$" Let X be periodic with period $d \geq 2$. If $n \in \mathbb{N}$ is not a multiple of d, then $p^n(x, x) = 0$. Hence, for $A = B = \{X_0 = x\}$,

$$\begin{aligned}\liminf_{n\to\infty} \mathbf{P}_\pi[X_0 = x, X_n = x] &= \liminf_{n\to\infty} \pi\big(\{x\}\big) p^n(x, x)\\ &= 0 \neq \pi\big(\{x\}\big)^2 = \mathbf{P}_\pi[X_0 = x]^2.\end{aligned}$$

Thus X is not mixing.

"$\Longleftarrow$" Let X be aperiodic. In order to simplify the notation, we may assume that X is the canonical process on $E^{\mathbb{N}_0}$. Let $A, B \subset \Omega = E^{\mathbb{N}_0}$ be measurable. For every $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ and a $\tilde{A}^\varepsilon \in E^{\{0,\ldots,N\}}$ such that, letting $A^\varepsilon = \tilde{A}^\varepsilon \times E^{\{N+1,N+2,\ldots\}}$, we have $\mathbf{P}[A \triangle A^\varepsilon] < \varepsilon$. By the Markov property, for every

$n \geq N$,

$$\begin{aligned}\mathbf{P}_\pi\big[A^\varepsilon \cap \tau^{-n}(B)\big] &= \mathbf{P}_\pi\big[(X_0, \ldots, X_N) \in \tilde{A}^\varepsilon, (X_n, X_{n+1}, \ldots) \in B\big] \\ &= \sum_{x,y\in E} \mathbf{E}_\pi\big[\mathbb{1}_{A^\varepsilon}\mathbb{1}_{\{X_N=x\}}\mathbb{1}_{\{X_n=y\}}(X_n, X_{n+1}, \ldots) \in B\big] \\ &= \sum_{x,y\in E} \mathbf{E}_\pi[\mathbb{1}_{A^\varepsilon}\mathbb{1}_{\{X_N=x\}}]p^{n-N}(x,y)\mathbf{P}_y[B].\end{aligned}$$

By Theorem 18.13, we have $p^{n-N}(x,y) \overset{n\to\infty}{\longrightarrow} \pi(\{y\})$ for all $x, y \in E$. (For periodic X, this is false.) Dominated convergence thus yields

$$\begin{aligned}\lim_{n\to\infty} \mathbf{P}_\pi\big[A^\varepsilon \cap \tau^{-n}(B)\big] &= \sum_{x,y\in E} \mathbf{E}_\pi[\mathbb{1}_{A^\varepsilon}\mathbb{1}_{\{X_N=x\}}]\pi\big(\{y\}\big)\mathbf{P}_y[B] \\ &= \mathbf{P}_\pi\big[A^\varepsilon\big]\mathbf{P}_\pi[B].\end{aligned}$$

Since $|\mathbf{P}_\pi[A^\varepsilon \cap \tau^{-n}(B)] - \mathbf{P}[A \cap \tau^{-n}(B)]| < \varepsilon$, the statement follows by letting $\varepsilon \to 0$. □

Exercise 20.5.1 Show that "strongly mixing" implies "weakly mixing", which in turn implies "ergodic". Give an example of a measure-preserving dynamical system that is ergodic but not weakly mixing.

20.6 Entropy

The entropy $H(\mathbf{P})$ of a probability distribution $\mathbf{P}$ (see Definition 5.25) measures the amount of randomness in this distribution. In fact, the entropy of a delta distribution is zero and for a distribution on n points, the maximal entropy is achieved by the uniform distribution and equals $\log(n)$ (see Exercise 5.3.3). It is natural to use the entropy in order to quantify also the randomness of a dynamical system.

First we consider the situation of a simple shift: Let $\Omega = E^{\mathbb{N}_0}$, where E is a finite set equipped with the product σ-algebra $\mathcal{A} = (2^E)^{\otimes\mathbb{N}_0}$. Let τ be the shift on Ω and let $\mathbf{P}$ be an invariant probability measure. For $n \in \mathbb{N}$, denote by P_n the projection of $\mathbf{P}$ on $E^n = E^{\{0,\ldots,n-1\}}$; that is,

$$P_n\big(\{(e_0, \ldots, e_{n-1})\}\big) = \mathbf{P}\big[\{e_0\} \times \ldots \times \{e_{n-1}\} \times E^{\{n,n+1,\ldots\}}\big].$$

Denote by h_n the entropy of P_n. By Exercise 5.3.4, the entropy is subadditive:

$$h_{m+n} \leq h_m + h_n \quad \text{for } m, n \in \mathbb{N}.$$

Hence the following limit exists (see Exercise 20.6.2)

$$h := h(\mathbf{P}, \tau) := \lim_{n\to\infty} \frac{1}{n} h_n = \inf_{n\in\mathbb{N}} \frac{1}{n} h_n.$$

Definition 20.30 (Entropy of the simple shift) $h(\mathbf{P}, \tau)$ is called the *entropy* of the dynamical system $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$.

Example 20.31 Assume that $\mathbf{P}$ is a product measure with marginals π on E. Then

$$h = H(\pi) = -\sum_{e \in E} \pi\big(\{e\}\big) \log\big(\pi\big(\{e\}\big)\big).$$

◊

Example 20.32 (Markov chain) Let $(X_n)_{n\in\mathbb{N}_0}$ be a Markov chain on E with transition matrix P and stationary distribution π. Let $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ be the corresponding dynamical system. For $x = (x_0, \ldots, x_{n-1})$ and $0 \le k < n-1$, let

$$p(k, x) = \pi\big(\{x_k\}\big) P(x_k, x_{k+1}) \ldots P(x_{n-2}, x_{n-1}).$$

Then the entropy of P_n is (using stationarity of π in the third line)

$$\begin{aligned} H(P_n) &= -\sum_{x_0,\ldots,x_{n-1}\in E} p(0,x) \log\big(p(0,x)\big) \\ &= -\sum_{x_0,\ldots,x_{n-1}\in E} p(0,x) \left[\log\big(\pi\big(\{x_0\}\big)\big) + \sum_{k=0}^{n-2} \log\big(P(x_k, x_{k+1})\big) \right] \\ &= H(\pi) - \sum_{k=0}^{n-2} \sum_{x_k,\ldots,x_{n-1}} p(k,x) \log\big(P(x_k, x_{k+1})\big) \\ &= H(\pi) - (n-1) \sum_{x_0,x_1\in E} \pi\big(\{x_0\}\big) P(x_0,x_1) \log\big(P(x_0,x_1)\big). \end{aligned}$$

We infer that the entropy of the dynamical system is

$$h(\mathbf{P}, \tau) = -\sum_{x,y\in E} \pi\big(\{x\}\big) P(x,y) \log\big(P(x,y)\big). \tag{20.9}$$

◊

Example 20.33 (Integer rotation) Consider the rotation of Example 20.8. Let $n \in \mathbb{N} \setminus \{1\}$, $E = \mathbb{Z}/(n)$ and let $\mathbf{P}$ be the uniform distribution on Ω. Let $r \in \{1, \ldots, n\}$ and

$$\tau : \Omega \to \Omega, \quad x \mapsto x + r \ (\mathrm{mod}\ n).$$

Clearly, $\tau^{(n)}$ is the identity map, hence $h_n = h_{2n} = \ldots$ and thus $h(\mathbf{P}, \tau) = 0$. ◊

We now come to the situation of the general dynamical system. Let $\mathcal{P}$ be a finite measurable partition of Ω; that is, $\mathcal{P} = \{A_1, \ldots, A_k\}$ for certain pairwise disjoint non-empty sets $A_1, \ldots, A_k \in \mathcal{A}$ with $\Omega = A_1 \cup \ldots \cup A_k$. Denote by $\mathcal{P}_n$ the partition that is generated by the sets $\bigcap_{l=0}^{n-1} \tau^{-l}(A_{i_l})$, $i_1, \ldots, i_n \in \{1, \ldots, k\}$. We define

$$h_n(\mathbf{P}, \tau; \mathcal{P}) = -\sum_{A\in\mathcal{P}_n} \mathbf{P}[A] \log\big(\mathbf{P}[A]\big).$$

Similarly as in the simple shift case, we obtain the subadditivity of (h_n) and thus the existence of

$$h(\mathbf{P}, \tau; \mathcal{P}) := \lim_{n\to\infty} \frac{1}{n} h_n(\mathbf{P}, \tau; \mathcal{P}) = \inf_{n\in\mathbb{N}} \frac{1}{n} h_n(\mathbf{P}, \tau; \mathcal{P}).$$

Definition 20.34 (Kolmogorov–Sinai entropy) The *entropy* of a (general) measure-preserving dynamical system $(\Omega, \mathcal{A}, \mathbf{P}, \tau)$ is

$$h(\mathbf{P}, \tau) = \sup_{\mathcal{P}} h(\mathbf{P}, \tau; \mathcal{P}),$$

where the supremum is taken over all finite measurable partitions of Ω.

Theorem 20.35 (Kolmogorov–Sinai) *Let $\mathcal{P}$ be a generator of $\mathcal{A}$; that is $\mathcal{A} = \sigma(\bigcup_{n\in\mathbb{N}_0} \tau^{-n}(\mathcal{P}))$. Then*

$$h(\mathbf{P}, \tau) = h(\mathbf{P}, \tau; \mathcal{P}).$$

Proof See, e.g., [88, Theorem 3.2.18], [167, Theorem 4.17] or [155]. □

The Kolmogorov–Sinai theorem shows that the entropy that was introduced in Definition 20.30 for simple shifts coincides with the entropy of Definition 20.34; simply take $\mathcal{P} = \{\{e\} \times E^{\mathbb{N}}, e \in E\}$ which generates the product σ-algebra on $\Omega = E^{\mathbb{N}_0}$.

Example 20.36 (Rotation) We come back to the rotation of Example 20.9. Let $\Omega = [0, 1)$, $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathbf{P} = \lambda$ the Lebesgue measure, $r \in (0, 1)$ and $\tau_r(x) = x + r \pmod 1$.

First assume that r is rational. Let $\mathcal{P}$ be an arbitrary finite measurable partition of Ω. Choose $n \in \mathbb{N}$ such that $rn \in \mathbb{N}_0$. As in Example 20.33 we obtain $h_n(\mathbf{P}, \tau_r; \mathcal{P}) = h_{kn}(\mathbf{P}, \tau_r; \mathcal{P})$ for all $k \in \mathbb{N}$, hence $h(\mathbf{P}, \tau_r, \mathcal{P}) = 0$. Concluding, we get $h(\mathbf{P}, \tau_r) = 0$.

Now assume that r is irrational. Choose the partition $\mathcal{P} = \{[0, 1/2), [1/2, 1)\}$. As r is irrational, it is easy to see that $\mathcal{A}$ is generated by $\bigcup_{n\in\mathbb{N}_0} \tau_r^{-n}(\mathcal{P})$. Hence $h(\mathbf{P}, \tau_r) = h(\mathbf{P}, \tau_r, \mathcal{P})$. In order to compute the latter quantity, we first determine the cardinality $\#\mathcal{P}_n$. To this end, consider the map

$$\begin{aligned} \phi_n &: [0, 1) \to \{0, 1\}^n \\ &\qquad x \mapsto \big(\mathbb{1}_{[1/2,1)}(x), \mathbb{1}_{[1/2,1)}\big(\tau_r(x)\big), \ldots, \mathbb{1}_{[1/2,1)}\big(\tau_r^{n-1}(x)\big)\big). \end{aligned}$$

Clearly, we have $\#\phi_n([0, 1)) = \#\mathcal{P}_n$. As $x \in [0, 1)$ increases, each coordinate $\mathbb{1}_{[1/2,1)}(\tau_r^k(x))$, $k = 1, \ldots, n-1$, changes its value exactly twice. Only $\mathbb{1}_{[1/2,1)}(x)$ changes the value exactly once. Summing up, we get $\#\phi_n([0, 1)) \le 2n$. The maximal entropy of a probability measure on N points is achieved by the uniform distribution and is $\log(N)$. Consequently, $h_n(\mathbf{P}, \tau_r; \mathcal{P}) \le \log(2n)$. We conclude that

$$h(\mathbf{P}, \tau_r) = h(\mathbf{P}, \tau_r; \mathcal{P}) = 0.$$ ◊

Exercise 20.6.1 Let $\Omega = [0, 1)$ and $\tau : x \mapsto 2x \pmod 1$. Let $\mathbf{P}$ be the Lebesgue measure on Ω. Determine $h(\mathbf{P}, \tau)$.

Exercise 20.6.2 Let $(a_n)_{n\in\mathbb{N}}$ be a sequence on nonnegative numbers. The sequence is called *subadditive*, if $a_{m+n} \le a_m + a_n$ for all $m, n \in \mathbb{N}$. Show that the limit $\lim_{n\to\infty} a_n/n$ exists and that

$$\lim_{n\to\infty} \frac{1}{n} a_n = \inf_{n\in\mathbb{N}} \frac{1}{n} a_n.$$